Sample Size Calculations for a Difference of Means

October 11, 2019

Computing Power For Two-Sample Tests

We have a clinical trial with 100 patients in each group and previous studies to suggest a standard deviation of 12.

We want to test

- $H_0 : \mu_{trt} - \mu_{ctrl} = 0$
- $H_A : \mu_{trt} - \mu_{ctrl} \neq 0$

Computing Power For Two-Sample Tests

- Last class, we found the standard error estimate $SE = 1.70$.
- This corresponds to rejecting if

$$\bar{x}_{trt} - \bar{x}_{ctrl} < -3.33$$

or

$$\bar{x}_{trt} - \bar{x}_{ctrl} > 3.33$$

Let's redraw the null and sampling distributions for the case where the truth is

$$\mu_{trt} - \mu_{ctrl} = -3$$

Determining Sample Size

With a sample size of 100...

- We can only detect an effect size of 3 mmHg with probability 0.42.
- This means that the probability of Type II error is 0.58!
- For drug development, Type II error could cost a company several hundred million dollars.

Determining Sample Size

It is of interest to determine an appropriate sample size. This will allow us to...

- be confident that we are detecting any important effects.
- possibly save money in the long-run.
- Balance the probability of Type I and Type II error.

Example

Suppose now that we recruit 500 patients for each group.

- Determine the new estimate for standard error.

Example

Suppose now that we recruit 500 patients for each group.

- Determine the new estimate for standard error.
- Identify the rejection regions.

Example

Suppose now that we recruit 500 patients for each group.

- Determine the new estimate for standard error.
- Identify the rejection regions.
- Sketch out the null and alternative distributions when $\mu_{trt} - \mu_{ctrl} = -3$.

Example

Suppose now that we recruit 500 patients for each group.

- Determine the new estimate for standard error.
- Identify the rejection regions.
- Sketch out the null and alternative distributions when $\mu_{trt} - \mu_{ctrl} = -3$.
- Compute the probability of accurately rejecting H_0.

Determining Sample Size

We need to find a better balance!

- We can use sample size to find power (we could try a bunch of different numbers)
- Or we can flip the problem and use power to find sample size.
  - We usually use a power around 80% or 90%.

Example

What sample size do we need to get a power of 80%?

1. Start by assuming normal distribution.
   - We will make a correction if later we find $n < 30$.

Example

What sample size do we need to get a power of 80%?

1. Start by assuming normal distribution.
   - We will make a correction if later we find $n < 30$.
2. Determine the minimum effect size you want to be able to detect.

Example

What sample size do we need to get a power of 80%?

1. Start by assuming normal distribution.
   - We will make a correction if later we find $n < 30$.
2. Determine the minimum effect size you want to be able to detect.
3. Find the Z-score corresponding to a lower tail of 0.80.
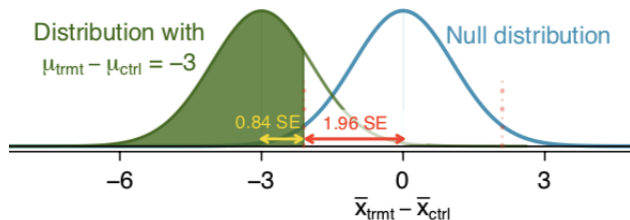
4. Find the number of standard deviations between the centers of the two distributions.

Example

- The distance between the centers of the two distributions is the minimum effect size of interest.
  - we set this when we chose $\mu_{trt} - \mu_{ctrl} = -3$.
- Let's try setting this equal to the distance in standard errors.
  - Can we us this to find n?

Example

Suppose the targeted power was 90% and we were using $\alpha = 0.01$.

1. How many standard errors should separate the centers of the null and alternative distribution?

2. What sample size would we need?
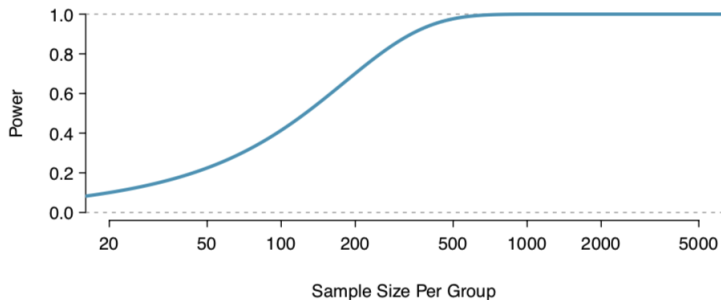
Power Considerations

Targeted power will depend on cost:

- Cost to patients (risk).
- Cost of enrolling study participants.
- Cost of failing to detect an effect of interest.

If risk and monetary costs are low it makes sense to increase power even if failing to detect an effect of interest isn't a big deal.

Power Considerations

However...

there's often a ceiling!

Example

A large farm wants to try out a new type of fertilizer to evaluate whether it will improve the farm's corn production. The land is broken into plots that produce an average of 1215 pounds of corn with a standard deviation of 94 pounds per plot. The owner is interested in detecting any average difference of at least 40 pounds per plot. How many plots of land would be needed for the experiment if the desired power level is 90%? Assume each plot of land gets treated with either the current fertilizer or the new fertilizer.